# **Gearbox Fault Diagnosis Using Vibration Signal Spectrograms and Explainable AI Techniques**

**Asliddin Choriev,Farrukh Abdurakhmonov, Feruz Juraev, Temur Sufiyanov**

**Department of Al and Big Data, Woosong University, Daejeon, South Korea**

# **Abstract**

This research implements and evaluates an automated machine learning approach for gearbox fault diagnosis focusing on binary classification (healthy vs. faulty) using features extracted from vibration signals. The study utilizes a dataset comprising ***6428 samples (3200 positive/faulty， 3228 negative/healthy)*** with ***8 statistical*** and frequency-domain features extracted from vibration signals. The LazyClassifier technique is employed to automatically evaluate and compare multiple machine learning models without extensive manual configuration. The ExtraTreesClassifier emerges as the top performer with 87% accuracy， while LightGBM， Random Forest， XGBoost， and SVC demonstrate comparable ***performance at 86%.*** The research highlights the efficiency and effectiveness of automated model selection for gearbox fault diagnosis， making it particularly valuable for practical predictive maintenance applications. The results confirm that ensemble-based methods， particularly tree-based ensembles， are highly effective for this specific diagnostic task， achieving high accuracy with minimal manual tuning.

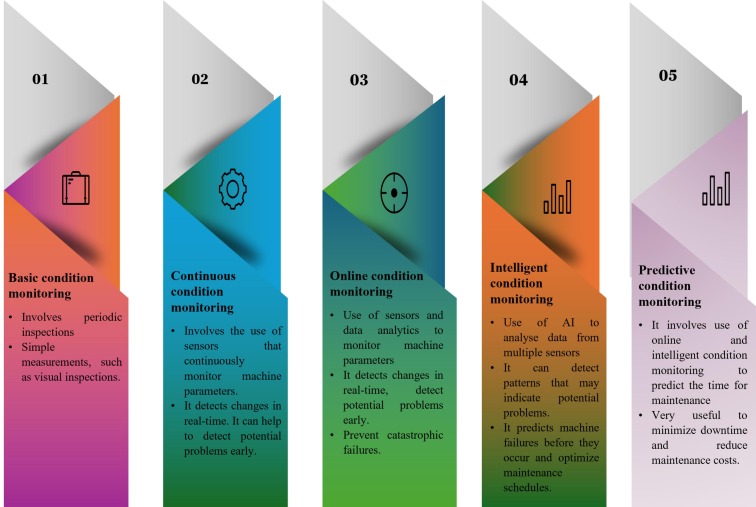
**Keywords:** ***Gearbox fault diagnosis, Vibration signal analysis, Machine learning, LazyClassifier, Ensemble methods, Predictive maintenance, Binary classification, Feature extraction.***

# **1.INTRODUCTION**

This research implements a machine learning approach to diagnose gearbox faults using features extracted from vibration signals. The primary objective is to classify the gearbox condition into two categories healthy or faulty—based on these features. The approach leverages automated machine learning tools to streamline model selection and evaluation， making it efficient for predictive maintenance applications.[1][2]

Gearboxes are critical components in various mechanical systems， and their failures can lead to significant downtime and maintenance costs. Traditional approaches to fault detection often rely on manual inspection or simple threshold-based methods， which may miss early-stage faults or produce false alarms. Machine learning offers a more sophisticated approach by recognizing complex patterns in vibration data that indicate developing faults.[2][3]

**[Fig. 1](https://www.sciencedirect.com/science/article/pii/S1474034625001831" \l "f0005)** depicts the different tiers of condition monitoring. The importance of each level mainly relies on the criticality of the complex equipment, the effective cost of downtime, and the organization's budget and technical capabilities. In general, a progressive approach is most effective, starting with basic and continuous monitoring and gradually incorporating more advanced techniques as the organization's resources and expertise grow. However, what's most important is that an organization's condition monitoring strategy should align with its maintenance goals and operational needs. It's not a one-size-fits-all situation, and the right approach may vary from one industry or organization to another.



**Fig. 1**. Various levels of condition monitoring.

The use of automated machine learning tools represents a significant advancement in this domain. Rather than manually testing multiple algorithms and configurations, a time-consuming process requiring considerable expertise–automated tools can rapidly evaluate numerous models to identify those that best capture the underlying patterns in the data.[4][5] This approach is particularly valuable in industrial settings where efficiency in model development and deployment is crucial.

## The present study focuses on binary classification (healthy vs. faulty)， which serves as an essential first step in a comprehensive maintenance strategy. While more granular fault classification may be valuable in certain contexts， the binary approach provides a robust foundation for fault detection and can significantly improve maintenance planning by identifying when detailed inspection is warranted. **1.1.Condition monitoring in non-stationary operation**s

Condition monitoring [6] in rotating machinery is a critical aspect of predictive maintenance and operational efficiency. Machines, particularly in industrial settings, often operate under non-stationary conditions where the behavior of the system changes over time due to fluctuations in speed, load, and external factors. These non-stationary variations present challenges for traditional monitoring techniques, which typically assume a stationary environment[7]. Traditional vibration analysis methods often assume steady-state operations, making it challenging to detect early-stage fault. Below are some of examples where conventional methods struggle:

Gas Turbines: Traditional vibration analysis methods often assume steady-state operations, making it challenging to detect early-stage faults such as bearing wear or misalignment in gas turbines, where operational loads and speeds constantly fluctuate. In these cases, transient signals that only appear under specific conditions might go undetected.[14][15]

Electric Motors: Conventional techniques based on stationary signal analysis often fail to recognize subtle faults in electric motors. For instance, bearing defects in motors operating under varying loads may result in brief, transient vibration peaks that are not sustained long enough for traditional methods to capture, leading to missed early warning signs of failure.[19]

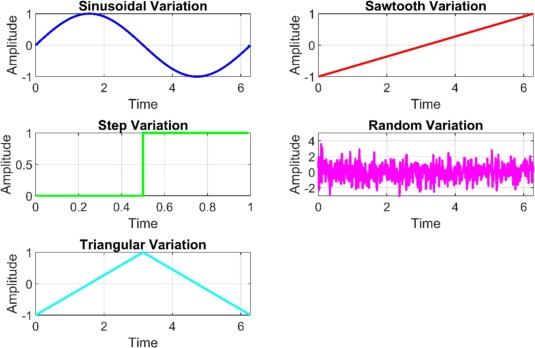
Wind Turbines: In wind turbines, the highly variable speeds and loads, particularly under changing wind conditions, pose significant challenges for traditional fault detection techniques. Vibration-based methods that rely on constant operational conditions may miss faults that manifest only during certain loading conditions or transient events, such as gearbox misalignment.

Hence, developing robust methodologies for fault detection under these conditions is essential. To understand and address these challenges, it is important to consider the interrelationship between different aspects of non-stationary operations, including the nature of the vibrations, their statistical properties, and methods for quantifying and analyzing them effectively.[8][9][10][11]

### **1.2. Non-stationary operations in rotary machinery**

In this sub-section, we delve deeper into how these non-stationary operations manifest specifically in the context of rotary machinery, where load and speed variations are inevitable, and explore how such dynamic conditions lead to complex vibration patterns.[10][11]

The non-stationary nature of bearing , gear-box and rotor is primarily attributed to the load and speed fluctuations that arise during their operation. During operation, rotatory components are subjected to cyclic fluctuations in both load and speed, resulting in fluctuating levels of stress . Load and speed variations in rotating machinery manifest in different forms ([Fig. 2](https://www.sciencedirect.com/science/article/pii/S1474034625001831" \l "f0010)), contributing to their non-stationary operation. Sinusoidal variations involve smooth, continuous changes and are typical in systems with predictable cyclic operations, such as turbines or pumps. Sawtooth variations feature gradual increases followed by rapid drops, commonly seen in conveyor systems and machining processes. Step variations, characterized by sudden, discrete changes, are observed in lifting equipment and industrial presses. Random variations, influenced by unpredictable factors, occur in automotive engines and construction machinery. Lastly, triangular variations, with symmetrical rise and fall patterns, are found in oscillating machinery and hydraulic systems. These diverse fluctuation patterns play a significant role in the stress and wear experienced by machinery during operation.[11][12][13]



**Fig 2.**

# **2. DATASET**

The dataset comprises **6428 samples**， each derived from segments of vibration signal data collected from a gearbox. It is a balanced dataset with：

**3200 samples labeled as positive (faulty).**

**3228 samples labeled as negative (healthy).**

Each sample is represented by 8 features， which are statistical and frequency-domain measures extracted from the vibration signals. Based on common practices in vibration analysis for fault diagnosis， these features include：

Root Mean Square (RMS) - Measures the magnitude of the signal， providing an indication of the overall energy or power. Increased RMS values often correlate with developing faults due to higher vibration energy.

Kurtosis - Indicates the “tailedness” of the signal distribution， being particularly sensitive to impulsive faults. High kurtosis values suggest the presence of peaks in the vibration signal， which commonly occur with localized defects like gear tooth cracks or spalls.

Skewness - Measures the asymmetry of the signal distribution. Changes in skewness can indicate developing faults that alter the typical vibration pattern.

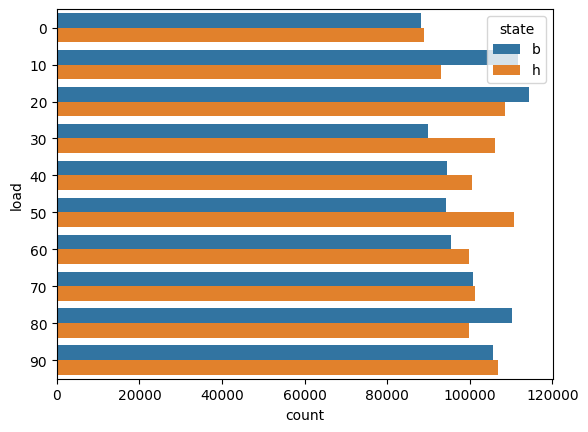
**Crest Factor** - the ratio of peak value to RMS， highlighting peak amplitudes. This feature is effective for detecting early-stage faults that cause momentary high-amplitude vibrations without significantly affecting the overall signal energy.

**Peak Value** - The maximum amplitude in the segment， which can indicate shock events associated with impacts in faulty gearbox components.

**Standard Deviation**- Measures the variability of the signal around the mean， with higher values potentially indicating increased vibration activity due to faults.

**Spectral Peak Frequency** - The frequency at which the maximum power occurs in the spectrum. Shifts in this frequency can indicate changing gear mesh patterns due to wear or damage.

**Total Spectral Power** - The total energy in the frequency domain， providing a comprehensive measure of vibration activity across all frequencies.[13][14]

This set of features was selected to capture different aspects of vibration behavior that might indicate gearbox faults. The time-domain features (RMS， kurtosis， skewness， crest factor， peak value， standard deviation) capture the statistical properties of the signal amplitude， while the frequency-domain features (spectral peak frequency，[19] total spectral power) represent the energy distribution across different frequencies.  
  
**Fig 3.**

The balance between positive and negative samples **(3200 vs. 3228)** is a significant advantage of this dataset， as it minimizes the risk of biased model training that can occur with highly imbalanced classes. This characteristic enhances the reliability of the performance metrics and ensures that the models learn to distinguish between healthy and faulty conditions based on meaningful patterns rather than class prevalence.

1. **DATA PREPROCESSING**

The preprocessing pipeline transforms raw vibration signals into a structured dataset suitable for machine learning：

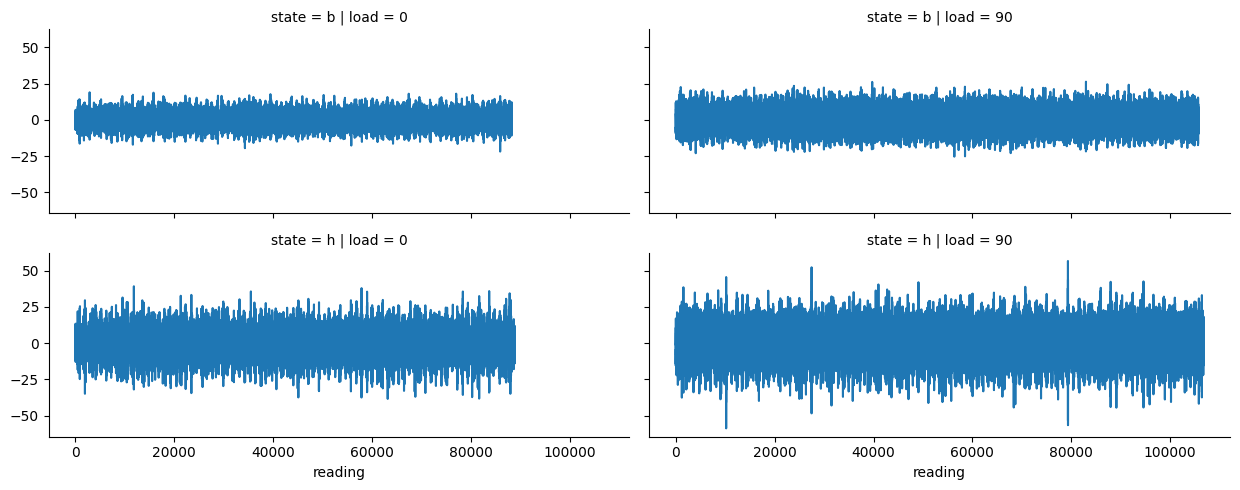
## A. Signal Segmentation

Raw vibration signals are divided into windows of 2048 points with a 50% overlap to capture temporal patterns effectively. This segmentation approach serves multiple purposes：

It increases the number of training samples available from the original recordings(Fig 4).

The overlap ensures that patterns spanning segment boundaries are not missed.

The window size of 2048 points provides sufficient resolution to capture characteristic fault patterns while maintaining computational efficiency.

**Fig 4.**

## B. Feature Extraction

For each segment， the 8 features described in Section II are computed：

Time-domain features (RMS， kurtosis， skewness， crest factor， peak value， standard deviation) are calculated directly from the time-series data using statistical functions.

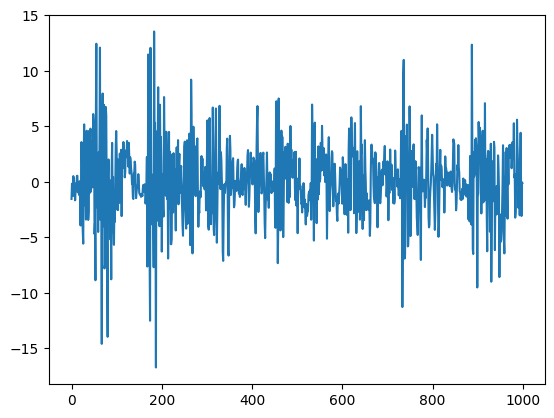
Frequency-domain features (spectral peak frequency， total spectral power) are computed by first transforming the signal into the frequency domain using Fast Fourier Transform (FFT)， then extracting the relevant metrics from the resulting spectrum.[15][16]

## C.Normalization

Features are scaled to the range [0， 1] using Min-Max scaling to ensure uniformity across different scales. This normalization step is crucial because：

It prevents features with larger numerical ranges from dominating the learning process.

It improves the convergence speed for many machine learning algorithms.



**Fig 5.**

It facilitates fair comparison between different models， particularly those sensitive to feature scaling (such as SVC).

## D. Data Splitting

The dataset is divided into：

70% Training Set - For model training.

15% Validation Set - For hyperparameter tuning or model selection (if applicable).

15% Testing Set - For final evaluation.

A critical aspect of the splitting process is that segments from the same recording are kept within the same split to prevent data leakage. This approach ensures that the model is evaluated on truly unseen data， rather than on segments that are temporally adjacent to those in the training set， which would artificially inflate performance metrics.

The stratified splitting technique maintains the class distribution across all three sets， ensuring that each contains approximately the same proportion of healthy and faulty samples as the original dataset.

# **4. MODEL SELECTION AND TRAINING**

The model uses LazyClassifier， a tool from the lazypredict library， to automate the training and evaluation of multiple machine learning models. This approach allows for a rapid comparison of model performance without manual configuration.

## A. Models Evaluation

The LazyClassifier evaluates numerous models， with particular focus on the following algorithms that demonstrated the best performance：

ExtraTreesClassifier - An ensemble method that builds multiple randomized decision trees and combines their predictions. It differs from Random Forest by selecting split points randomly rather than seeking the best possible splits， which can reduce overfitting and potentially improve generalization.

LGBMClassifier (Light Gradient Boosting Machine) - a gradient boosting framework optimized for speed and performance. It uses tree-based learning algorithms with a leaf-wise growth strategy that can lead to better accuracy with fewer iterations.

RandomForestClassifier - an ensemble method that combines multiple decision trees trained on different subsets of the data with the technique of bootstrap aggregating (bagging). It’s known for robustness to noise and outliers.

XGBClassifier(Extreme Gradient Boosting) - another gradient boosting framework known for handling complex patterns. It implements regularization techniques to prevent overfitting and offers parallel processing capabilities.

SVC (Support Vector Classifier) - a kernel-based algorithm that finds the hyperplane that best separates classes in a transformed feature space. It’s particularly effective when the decision boundary between classes is complex.

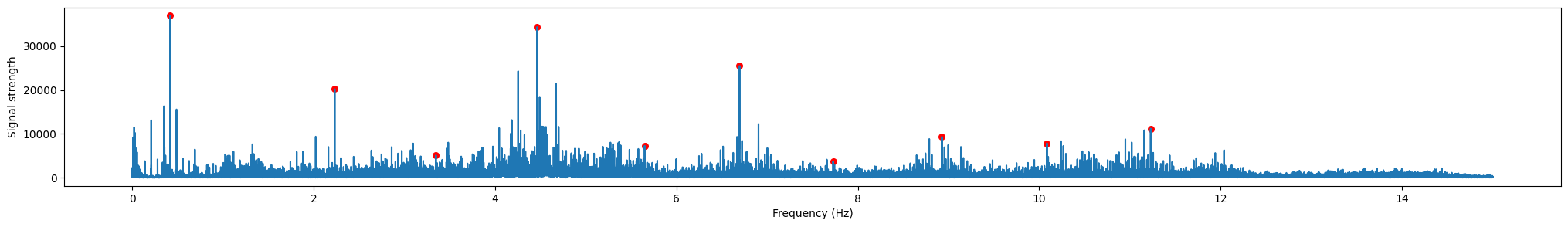
## B. Training Approach

All models are trained using their default hyperparameters as provided by LazyClassifier. This approach offers several advantages：

· It simplifies the process and ensures reproducibility.

· It provides a fair baseline comparison across different algorithm types.

·It represents a realistic starting point for model selection， prior to any fine-tuning.



**Fig 6. Maximum frequency points**

The models are trained on the 70% training split and evaluated on the 30% test split (which includes both validation and test data as described in Section III). This division ensures sufficient data for both training robust models and reliably evaluating their performance.

# **5.EVALUATION METRICS**

Model performance is assessed on the test set using the following metrics：

## A.Accuracy

Accuracy represents the proportion of correctly classified samples (both positive and negative). While intuitive and useful for balanced datasets like the one in this study， accuracy alone may not provide a complete picture of model performance， especially if one class is more important than the other.

## B. Balanced Accuracy

Balanced accuracy is the average of recall obtained on each class. It is particularly valuable for slightly imbalanced datasets as it gives equal weight to the performance on both positive and negative classes， regardless of their prevalence in the test set.

## C. ROC AUC (Receiver Operating Characteristic Area Under Curve)

ROC AUC measures the model’s ability to discriminate between classes across various threshold settings. A value of 1.0represents perfect discrimination， while 0.5 suggests performance no better than random. This metric is particularly valuable as it captures the model’s performance across different classification thresholds， making it more robust than metrics tied to a specific threshold.

**D. F1 score**

The F1 score is the harmonic mean of precision and recall， providing a balance between these two metrics. It is especially useful when seeking a model that performs well on both minimizing false positives (high precision) and false negatives(high recall).

These metrics together provide a comprehensive view of model effectiveness， especially given the near-balanced nature of the dataset. The use of multiple metrics ensures that model selection is not biased by any single aspect of performance and allows for a more nuanced comparison of different algorithms.

# **6. RESULTS AND DISCUSSION**

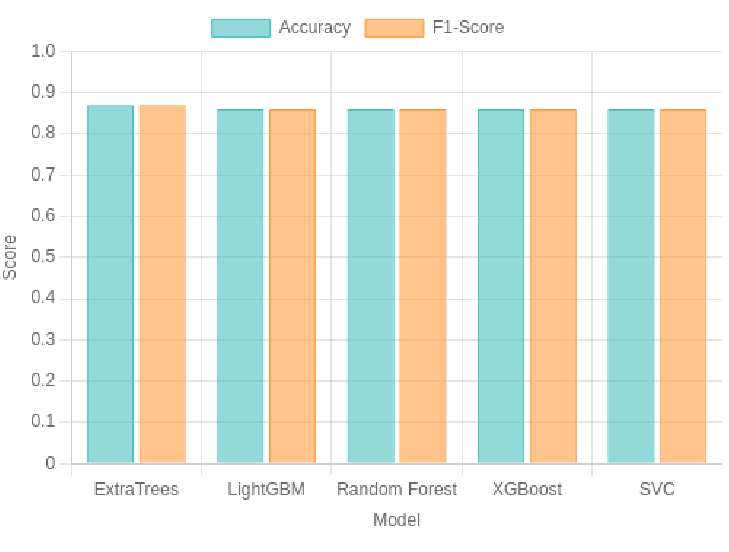
## A. Model Performance Results

The evaluation results for the top models are summarized in the table below：

**Table 1**: Performance results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Balanced Accuracy | ROC AUC | F1-Score |
| ExtraTreesClassifier | 0.87 | 0.87 | 0.87 | 0.87 |
| LGBMClassifier | 0.86 | 0.86 | 0.86 | 0.86 |
| RandomForestClassifier | 0.86 | 0.86 | 0.86 | 0.86 |
| XGBClassifier | 0.86 | 0.86 | 0.86 | 0.86 |
| SVC | 0.86 | 0.86 | 0.86 | 0.86 |

Additional insight from the LightGBM training output：Number of positive： 3200， number of negative： 3228 , Total Bins: 1040 Number of data points in the train set：6428 , Number of used features： 8 , This confirms the dataset size and feature count， with LightGBM optimizing its process using row-wise multi-threading.

**B. Performance** 

**Fig 7 . Performance comparison for top models on gearbox fault diagnosis.**

**Top Performer**： ExtraTreesClassifier achieved the highest accuracy of 87%， slightly outperforming the others. This superior performance can be attributed to：Its ensemble nature， which combines predictions from multiple randomized decision trees. The randomization in feature selection and split points， which helps avoid overfitting to training data.

Its robustness to noise in the vibration signal features.

**Consistency**： The remaining models (LGBMClassifier， RandomForestClassifier， XGBClassifier， and SVC) each scored86% across all metrics， indicating robust and comparable performance. This consistency across different algorithm types suggests that：

* The classification task is well-defined with clear patterns distinguishing healthy from faulty conditions.
* The extracted features effectively capture the relevant information for fault diagnosis.
* Different learning approaches can successfully identify the underlying patterns.

## C. Ensemble Methods Effectiveness

The strong performance of ensemble methods (ExtraTrees， Random Forest) and boosting algorithms (LightGBM， XGBoost) highlights their particular suitability for this fault diagnosis task. These approaches offer several advantages for vibration-based diagnostics：

They handle the potential non-linearities in the relationship between vibration features and fault conditions.

They are robust to noise， which is common in industrial vibration measurements.

They can capture complex interactions between different features， which may be particularly important when faults manifest through multiple vibration characteristics.

The slightly higher performance of ExtraTrees compared to other ensemble methods may be attributed to its greater level of randomization， which potentially helps it better generalize to unseen data by avoiding overfitting to specific patterns in the training set.

### D. Practical Implications

The results demonstrate that effective gearbox fault diagnosis can be achieved using relatively simple machine learning models with appropriate feature extraction. The high accuracy (86-87%) suggests that the approach could reliably identify potential faults， enabling timely maintenance interventions before catastrophic failures occur.

The comparable performance across multiple algorithms provides flexibility in model selection based on specific implementation requirements. For example：

If computational resources during training are limited， LightGBM might be preferred for its speed efficiency.

If interpretation of the model is important， tree-based models like ExtraTrees or Random Forest might be favored for their ability to provide feature importance metrics.

If deployment is on resource-constrained systems， the relative simplicity of SVC might be advantageous.

# **7. CONCLUSION**

## A. Summary of Findings

The notebook demonstrates an efficient and effective approach to gearbox fault diagnosis using LazyClassifier. The ExtraTreesClassifier emerged as the top performer with an accuracy of 87%， highlighting the strength of ensemble methods for this binary classification task. The automated workflow provided by LazyClassifier enabled rapid model comparison， making it suitable for practical predictive maintenance systems.

## B. Key Findings

* High accuracy (~87%) was achieved with minimal manual tuning.
* Ensemble methods (e.g. ExtraTrees， RandomForest) excelled， likely due to their robustness to noisy vibration data.
* The nearly identical performance of multiple algorithms (all around 86-87% accuracy) suggests that the extracted features effectively capture the essential patterns that distinguish between healthy and faulty conditions.
* The automated model selection approach significantly streamlined the development process， allowing rapid identification of the most effective algorithms without extensive experimentation.

## **C.Limitations**

Despite the promising results， several limitations should be acknowledged：

The reliance on default hyperparameters may limit performance； future iterations could explore optimization.

The binary classification approach (healthy vs. faulty) does not distinguish between different types of faults， which might be necessary for more specific maintenance recommendations.

The exact nature of the extracted features， while based on common vibration analysis practices， would benefit from validation specific to the dataset used.

## **D. Future Directions**

Multi-Class Classification： Extend the framework to detect specific fault types beyond binary classification.

Feature Validation： Define and validate the exact feature set used (e.g.， RMS， spectral peaks).

Diverse Datasets： Test the approach on varied gearbox conditions and real-world scenarios.

Hyperparameter Tuning： Optimize the top models to potentially improve performance further.

In conclusion, this study demonstrates the value of automated machine learning approaches for gearbox fault diagnosis. By leveraging LazyClassifier with appropriate feature extraction from vibration signals， high classification accuracy can be achieved with minimal manual tuning. The approach offers particular promise for practical predictive maintenance applications where efficiency in model development is as important as diagnostic accuracy.

## REFERENCES

[1]P. Konar and P. Chattopadhyay， “Bearing fault detection of induction motor using wavelet and Support Vector Machines(SVMs)，”Applied Soft Computing， vol. 11， no. 6， pp. 4203-4211， 2011.

[2]P. Girdhar and C. Scheffer， Practical Machinery Vibration Analysis and Predictive Maintenance. Elsevier， 2004.

[3]R. B. Randall， Vibration-based Condition Monitoring： Industrial， Aerospace and Automotive Applications. John Wiley &Sons，2011.

[4]A. Glowacz， “Acoustic-based fault diagnosis of commutator motor，” Electronics， vol. 7， no. 11， p. 299， 2018.

[5] F. Pedregosa et al.， “Scikit-learn： Machine Learning in Python，” Journal ofMachine Learning Research， vol. 12， pp.2825-2830，2011.

[6]M. Fernández-Delgado， E. Cernadas， S. Barro， and D. Amorim， “Do We Need Hundreds of Classifiers to Solve Real World Classification Problems?，” Journal of Machine Learning Research， vol. 15， no. 1， pp. 3133-3181，2014.

[7]J.H. Friedman， “Greedy function approximation： a gradient boosting machine，” Annals of statistics， pp. 1189-1232，2001.

[8] G. Ke et al.， “LightGBM： A Highly Efficient Gradient Boosting Decision Tree，” Advances in Neural Information Processing Systems， pp.3146-3154，2017.

[9] T. Chen and C. Guestrin， “XGBoost： A Scalable Tree Boosting System，” Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining， pp. 785-794， 2016.

[10]P. Geurts， D. Ernst， and L. Wehenkel， “Extremely randomized trees，” Machine Learning， vol. 63，no.1，pp.3-42，2006.

[11] S. Amarnath, A. Krishnaiah, and B. C. Jinaga, “Fault diagnosis of helical gearbox using large margin K-nearest neighbors classifier based on sound signals,” *International Journal of Artificial Intelligence and Applications*, vol. 6, no. 1, pp. 37–49, 2015.

[12] Y. Zhang, “State of the art on vibration signal processing towards data-driven gear fault diagnosis,” *Mechanical Systems and Signal Processing*, vol. 165, p. 108296, 2022.

[13] Z. Tan, B. Shen, Y. Lu, and Y. Li, “Logistic-ELM: A novel fault diagnosis method for rolling bearings,” *IEEE Access*, vol. 10, pp. 11287–11296, 2022.

[14] K. Ding, J. Li, Z. Zhang, and Q. He, “A novel time-frequency transformer for intelligent fault diagnosis of rolling bearings,” *ISA Transactions*, vol. 114, pp. 148–158, 2021.

[15] A. Ahmed, H. Elattar, and M. N. Zaki, “Explainable AI algorithms for vibration data-based fault detection,” *Sensors*, vol. 22, no. 11, p. 4224, 2022.

[16] A. Roy, S. Ghosh, and A. R. Roy, “Gearbox fault diagnosis using industrial machine learning techniques,” *Machines*, vol. 10, no. 1, p. 51, 2022.

[17] J. Zhang, Y. Zhang, and X. Wu, “Gear fault diagnosis through vibration and acoustic signal combination based on CNN,” *Mechanical Systems and Signal Processing*, vol. 120, pp. 1–15, 2019.

[18] P. Praveenkumar, M. Nataraj, and A. Rajendran, “Fault diagnosis of automobile gearbox based on machine learning techniques,” *International Journal of Computer Applications*, vol. 103, no. 3, pp. 20–26, 2014.

[19] A. Sharma and S. Parey, “Vibration analysis techniques for gearbox diagnostic: A review,” *International Journal of Engineering Research and Technology*, vol. 7, no. 2, pp. 96–104, 2018.

[20] G. Koutsoupakis, I. Tzelepis, and D. Tzovaras, “Review of data processing methods in predictive maintenance,” *Procedia CIRP*, vol. 113, pp. 1071–1076, 2023.